Non parametric estimation of hazard rate in presence of censoring

Sandra Plancade

Metrika manuscript No.
(will be inserted by the editor)

Non parametric estimation of hazard rate in presence of censoring

Sandra Plancade

Received: date / Accepted: date

Abstract This note presents an estimator of the hazard rate function based on right censored data. A collection of estimators are built from a regression-type contrast, in a general collection of linear models. Then, a penalised model selection procedure gives an estimator which satisfies an oracle inequality. In particular, we can prove that it is adaptive in the minimax sense on Hölder spaces.

Keywords Hazard · Censoring · Model selection · adaptivity

1 Introduction

In medical follow-up and other subjects, the observation of a variable of interest, for example the lifetime of an individual, can be right censored. This means that we only observe the minimum of the lifetime and a variable called censoring time (for example the time when a patient leaves the medical program), which is supposed independent of the lifetime. We also observe if this minimum corresponds to the variable of interest or to the censoring time. More precisely, we consider a sample $(X_i)_{i=1,\dots,n}$ of non-negative variables, and a sample $(C_i)_{i=1,\dots,n}$ of non-negative censoring times. Then we observe a sample $(Y_i, \delta_i)_{i=1,\dots n}$ with:

$$Y_i = \min(X_i, C_i), \quad \delta_i = 1_{X_i \leq C_i} \tag{1}$$

A function of interest in such a study is the hazard rate function of X, which represents the risk of death at a time x knowing that the patient is alive until x. If we denote by $f_X(x)$ and $\overline{F}_X(x) = P[X_1 \geq x]$ the density and the survival function of X, we have:

$$h(x) = \frac{f_X(x)}{\overline{F}_X(x)} \tag{2}$$

A lot of papers are devoted to hazard rate estimation, among which two general methods can be drawn in the non parametric context that we only consider.

The first one consists in estimating h by a quotient of two estimators. The most obvious is $\widehat{f}_X/\widehat{F}_X$ where $\widehat{f}_X$ and $\widehat{F}_X$ are estimators of f_X and $\overline{F}_X$. In general, $\overline{F}_X$ is replaced by the well known Kaplan Meier estimator of $\overline{F}_X$ (Kaplan and Meier (1958)). Another decomposition of h is

$$h = \frac{f_X \overline{F}_C}{\overline{F}_Y} \tag{3}$$

S. Plancade
MAP 5 UMR 8145
Universit Paris Descartes
45 rue des Saints Pères 75006 Paris, France.
Tel.: +33142863357
Fax: +33142864144
E-mail: sandra.plancade@math-info.univ-paris5.fr

The function $\psi(x) = f_X(x)\overline{F}_C(x)$, called the subdensity of X, corresponds heuristically to the "density" of the observed variables X_i, in the sense that for every function $t : \mathbb{R}^+ \to \mathbb{R}$ such that $t(0) = 0$:

$$\mathbb{E}[t(\delta_i Y_i)] = \mathbb{E}[\delta_i t(X_i)] = \int t(x)\psi(x)dx$$

As the (δ_i, X_i) are directly measured, ψ is easier to estimate than f_X. Similarly, $\overline{F}_Y$ is easier to estimate than $\overline{F}_X$. Indeed it can simply be replaced by the empirical survival function of the observed (Y_i). Patil (1993) proposes a kernel estimator of ψ with a bandwidth selection and gets an estimator of h via expression (3). Antoniadis et al (1999) use a wavelet decomposition but their estimator is not really adaptive as the optimal resolution of the wavelets depends on the regularity of f_X. Comte and Brunel (2005) build a projection estimator of ψ by model selection in more general bases, and obtain an adaptive estimator.

Other estimators of h are based on the cumulative hazard $H(x) = -\log(\overline{F}_X(x))$. One of the most frequently used estimator of H is the Nelson-Aalen estimator (Nelson (1972)). Obviously, we have:

$$h(x) = H'(x) \tag{4}$$

Yandell (1983) and Tanner and Wong (1983) build an estimator of h by differentiating the Nelson-Aalen estimator of H with a delta-sequence method, and Muller and Wang (1994) introduce a variable bandwidth. Comte and Brunel (2008) propose a projection type estimator based on a approximation of cumulative hazard function. The method is very different from the one presented here, but leads also to an adaptive estimation procedure.

Let us mention also the estimator of Reynaud-Bouret (2006) built by model selection in a set of random models, which is adaptive on Hölder spaces with regularity smaller than 1.

The present note describes a regression type strategy, in a different spirit from other procedures. It leads to an adaptive estimator for the integrated squared risk on a set $[0, \tau]$ such that $P(Y \geq \tau)$ is positive. The proofs are self contained (apart from the well-known Talagrand Inequality), and the key point is that the reference norm for the risk is chosen to be well suited to the problem.

The plan of the paper is the following. Section 2 presents the framework, and the main assumptions. The estimation procedure is described in Section 3, as well as the main result. But the estimator built in Section 3 brings into play unknown quantities, which are estimated in Section 4. The proofs are gathered in Sections 5 and 6. Section 7 recalls classical deviation inequalities for empirical processes.

2 Presentation of the framework, assumptions and notations

2.1 Framework

We consider a sample $(X_1, \ldots, X_n)$ of i.i.d. (independent identically distributed) non negative random variables with common survival function $\overline{F}_X(x) := P[X_1 \geq x]$ and density f_X, and a sample $(C_1, \ldots, C_n)$ of i.i.d. non negative random variables with common survival function $\overline{F}_C$. We suppose that the (C_i) are independent of the (X_i) The variables of interest are the (X_i), but we only observe the sample $((Y_1, \delta_1), \ldots, (Y_n, \delta_n))$ defined in (1). The aim of this paper is to build an estimator of the hazard rate of X_1 given by (2), on a compact interval A on which $\overline{F}_Y = \overline{F}_C\overline{F}_X$ is lower bounded by a positive number. Theoretically, A is a known compact interval independent of the data, even if practically it is chosen by looking at the data. Moreover, we take $A = [0, 1]$ for simplicity, but the results can be adapted to any compact interval A by rescaling the data. More precisely, we consider the following assumptions:

$\mathbf{A_{frame}}$: We suppose that $\overline{F}_Y$ is lower bounded on $A = [0, 1]$ by $\overline{F}_0 > 0$, and that h is upper bounded by $\|h\|_{\infty,A} := \sup_{x\in A} h(x) < \infty$.

2.2 Notations

We define the following scalar products and norms on $L^2(A)$. For every $s,t \in L^2(A)$:

$$\langle s,t\rangle = \int_A s(x)t(x)dx, \quad \|t\|^2 = \int_A t^2(x)dx$$

$$\langle s,t\rangle_{\overline{F}_Y} = \int_A s(x)t(x)\overline{F}_Y(x)dx, \quad \|t\|^2_{\overline{F}_Y} = \int_A t^2(x)\overline{F}_Y(x)dx$$

$$\langle s,t\rangle_n = \frac{1}{n}\sum_{i=1}^n \int_A s(x)t(x)1_{Y_i\geq x}dx, \quad \|t\|^2_n = \frac{1}{n}\sum_{i=1}^n \int_A t^2(x)1_{Y_i\geq x}dx$$

Let A be a square matrix, we denote by $Sp(A)$ the spectrum of A, that is the set of its eigenvalues.

Let β and L be positive numbers, and r the greatest integer smaller than β, we define the Hölder space $\mathcal{H}(\beta,L)$ on A:

$$\mathcal{H}(\beta,L) = \{f : A \to \mathbb{R}, |f^{(r)}(x) - f^{(r)}(y)| \leq L|x-y|^{\beta-r}, \forall x,y \in A\}$$

For every $x \in \mathbb{R}$, we denote by $E(x)$ the integer part of x, that is the greatest integer smaller than or equal to x.

All throughout the paper, C_i denotes a universal numerical constant, and C, C' denote constants which depend on the given parameters of the problem and may change from one line to another.

2.3 Collections of models

The estimators proposed in this paper are computed by model selection in a general collection of models $\mathcal{M}_n := \{S_m, m \in J_n\}$ where every model S_m is a finite dimensional linear subset of $L^2(A)$ with dimension D_m. We suppose that the collection $\mathcal{M}_n$ satisfies either Assumption $\mathbf{A^{(1)}_{mod}}$ or $\mathbf{A^{(2)}_{mod}}$.

$\mathbf{A^{(1)}_{mod}}$: The models S_m are nested, that is $J_n = \{1,\dots,N_n\}$ and:

$$S_1 \subset S_2 \subset \dots \subset S_{N_n} := S_n$$

Besides, there exists a constant K such that for every model S_m, and for every $(\phi_1^m,\dots,\phi_{D_m}^m)$ orthonormal basis of S_m for the $L^2(A)$-norm:

$$\sup_{x\in A} |\sum_{\lambda=1}^{D_m} (\phi_\lambda^m(x))^2| \leq K^2 D_m \tag{5}$$

Moreover, the maximum size of model satisfies: $N_n \leq n/\ln^2 n$.

$\mathbf{A^{(2)}_{mod}}$: There exists a linear subset S_n of $L^2(A)$ with dimension $N_n \leq n/\ln^2 n$ such that every model S_m is a subspace of S_n, and the global space S_n satisfies Property (5).

Remark 1 Obviously, Assumption $\mathbf{A^{(1)}_{mod}}$ is stronger than Assumption $\mathbf{A^{(2)}_{mod}}$. Thus, $\mathbf{A^{(2)}_{mod}}$ allows more irregular collections of models. Let us explain this notion on the example of histograms. Let I_n be the regular partition of $[0,1]$ of step $1/N_n$, and S_n the set of histograms on $[0,1]$ which are constant on I_n. Under Assumption $\mathbf{A^{(2)}_{mod}}$, the collection $\mathcal{M}_n$ can include any set of histograms S_m based on a partition I_m of $[0,1]$ composed of union of intervals from I_n. Whereas Assumption $\mathbf{A^{(1)}_{mod}}$ only allows diadic regular set of histograms, that is:

$$S_m = Vect\left\{1_{[\frac{j-1}{D_m},\frac{j}{D_m}[}, j = 1,\dots D_m\right\}, \text{ where } D_m = 2^{k_m}$$

for some $k_m = 1,\dots,k_n$ and $2^{k_n} = N_n$. We consider also a general assumption related to the maximum number of models for a given dimension.

$\mathbf{A^{(3)}_{mod}}$: For every $a > 0$, there exists a constant $A > 0$ such that, for every $n \in \mathbb{N}^*$,

$$\sum_{m\in J_n} \exp(-a\sqrt{D_m}) \leq A$$

3 Theoretical estimators

The estimators built in this section bring into play unknown quantities, that is why we call them theoretical estimators. We replace in turn these quantities by estimators in Section 4. In Section 3.1, we present a non adaptive procedure, to build an estimator $\widehat{h}_m$ of h on each model S_m. The model selection procedure is described in Section 3.2, with two different penalties corresponding to the two Assumptions $\mathbf{A^{(1)}_{mod}}$ and $\mathbf{A^{(2)}_{mod}}$. Section 3.4 presents the main result.

3.1 Non adaptive estimator

We consider the following contrast, for every $t \in L^2(A)$:

$$\gamma_n(t) = \|t\|_n^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(Y_i)$$

Let us justify this contrast. First,

$$h1_A = \arg\min_{t\in L^2(A)} \|t-h\|^2_{\overline{F}_Y} = \arg\min_{t\in L^2(A)} \|t\|^2_{\overline{F}_Y} - 2\langle t,h\rangle_{\overline{F}_Y}$$

Besides for every $t \in L^2(A)$, $\mathbb{E}[\|t\|_n^2] = \|t\|^2_{\overline{F}_Y}$. In addition, for every $i = 1,\dots,n$:

$$\begin{aligned}\mathbb{E}[\delta_i t(Y_i)] &= \mathbb{E}[\mathbb{E}\left[\delta_i t(Y_i)|X_i\right]] = \mathbb{E}\left[t(X_i)\mathbb{E}[1_{X_i\le C_i}|X_i]\right] \\ &= \mathbb{E}[t(X_i)\overline{F}_C(X_i)] = \int_A t(x)\overline{F}_C(x)f_X(x)dx \\ &= \int_A t(x)\overline{F}_C(x)\overline{F}_X(x)h(x)dx = \langle t,h\rangle_{\overline{F}_Y} \qquad (6)\end{aligned}$$

Thus $\mathbb{E}[\gamma_n(t)] = \|t\|_{\overline{F}_Y} - 2\langle t,h\rangle_{\overline{F}_Y}$ and $h1_A = \arg\min_{t\in L^2(A)}\mathbb{E}[\gamma_n(t)]$. This explains how $\gamma_n(t)$ is built.

For every model S_m, we define $\tilde{h}_m = \arg\min_{t\in S_m}\gamma_n(t)$. Let $(\phi_1^m,\dots,\phi_{D_m}^m)$ be an $L^2(A)$-orthonormal basis of S_m, then $\tilde{h}_m = \sum_{\lambda=1}^{D_m}\tilde{a}_\lambda\phi_\lambda^m$ with:

$$\frac{\partial\gamma_n(\sum_{\lambda=1}^{D_m}a_\lambda\phi_\lambda^m)}{\partial a_\lambda} = 0, \quad \forall\lambda = 1,\dots D_m$$

This is equivalent to:

$$\widehat{G}_m\tilde{A}_m = \widehat{V}_m$$

with $\tilde{A}_m = (\tilde{a}_1,\dots,\tilde{a}_{D_m})^t$ and:

$$\widehat{G}_m = (\langle\phi_\lambda^m,\phi_{\lambda'}^m\rangle_n)_{\lambda,\lambda'=1,\dots,D_m}, \quad \widehat{V}_m = \left(\frac{1}{n}\sum_{i=1}^n\delta_i\phi_\lambda^m(Y_i)\right)_{\lambda=1,\dots,D_m} \qquad (7)$$

where M^t denotes the transpose of M. But the matrix $\widehat{G}_m$ is not necessarily invertible. We consider a set on which $\widehat{G}_m$ is invertible, more precisely on which the spectrum of $\widehat{G}_m$ is lower bounded. First, we define the following set:

$$\Delta_1 = \left\{ \left|\frac{\|t\|_n^2}{\|t\|^2_{\overline{F}_Y}} - 1\right| < \frac{1}{4}, \forall t\in S_n\right\} \qquad (8)$$

Since $\|.\|_n$ is the empirical norm associated to $\|.\|_{\overline{F}_Y}$, $P[\Delta_1]$ is close to 1 (see Proposition 53). Let $t \in S_n$. On the set Δ_1 and under Assumption $\mathbf{A_{frame}}$, we have

$$\frac{3}{4}\overline{F}_0\|t\|^2 \le \frac{3}{4}\|t\|^2_{\overline{F}_Y} \le \|t\|_n^2 \qquad (9)$$

so that for every $\lambda \in Sp(\widehat{G}_m)$,

$$\frac{3}{4}\overline{F}_0 \leq \lambda. \tag{10}$$

Then we consider the following set, on which $\widehat{G}_m$ is invertible:

$$\Delta_2^{th} = \left\{ \min(Sp(\widehat{G}_m)) \geq \frac{3}{4}\overline{F}_0 \right\} \tag{11}$$

and it satisfies $\Delta_1 \subset \Delta_2^{th}$. Finally, we consider the estimator of h: $\widehat{h}_m = \sum_{\lambda=1}^{D_m} \widehat{a}_\lambda \phi_\lambda^m$ with

$$\widehat{A}_m = (\widehat{a}_1, \ldots, \widehat{a}_{D_m})^t = \begin{cases} \widehat{G}_m^{-1}\widehat{V}_m & \text{on } \Delta_2^{th} \\ 0 & \text{otherwise} \end{cases}$$

3.2 Adaptive estimators

By the non adaptive estimation procedure described above, we obtain a collection of estimators $\{\widehat{h}_m, m \in \mathcal{M}_n\}$, among which one is automatically selected by a penalised model selection procedure. We briefly present this strategy, developped by Birgé and Massart (1998). For every model S_m the risk of the estimator $\widehat{h}_m$ is split in two terms:

$$\mathbb{E}[\|\widehat{h}_m - h\|^2_{\overline{F}_Y} \leq 2\left(\|h - h_m\|^2_{\overline{F}_Y} + \mathbb{E}[\|\widehat{h}_m - h_m\|^2_{\overline{F}_Y}]\right)$$

where h_m is the $\|.\|_{\overline{F}_Y}$-projection of h on S_m The bias term $\|h - h_m\|^2_{\overline{F}_Y}$ decreases when the model S_m grows, whereas the term $\mathbb{E}[\|\widehat{h}_m - h_m\|^2_{\overline{F}_Y}]$ has the order D_m/n of a vairiance-type term, and increases with D_m. (Nevertheless in our case it is not exactly a variance, as $\mathbb{E}[\widehat{h}_m(x)] \neq h_m(x)$.) Thus, the best model would be the one which realises the better trade-off between bias and variance.

The basic outline of model selection is to estimate the bias-variance sum (possibly up to a constant independent of m) and to select the model which minimises this sum. Besides,

$$\|h - h_m\|^2_{\overline{F}_Y} = \|h_m\|^2_{\overline{F}_Y} - 2\langle h, h_m\rangle_{\overline{F}_Y} + \|h\|^2_{\overline{F}_Y}.$$

The term $\|h_m\|^2_{\overline{F}_Y} - 2\langle h, h_m\rangle_{\overline{F}_Y}$ is estimated by $\gamma_n(\widehat{h}_m)$ (see 6). The variance term $\mathbb{E}[\|\widehat{h}_m - h_m\|^2_{\overline{F}_Y}]$ is upper bounded by a deterministic term with order D_m/n, called the penalty. We do not explicitly prove this result here but a more general one (see Theorem 1 and Comment 1. hereafter).

We consider two penalties with order D_m/n, but with different constants.

$$pen_1^{th}(m) = \frac{BK^2}{\overline{F}_0}\frac{D_m}{n}, \quad pen_2^{th}(m) = B\|h\|_{\infty,A}\frac{D_m}{n} \tag{12}$$

with $B > 3$, and select the model

$$\widehat{m}_i = \arg\min_{m\in\mathcal{M}_n} \gamma_n(\widehat{h}_m) + pen_i^{th}(m)$$

for $i = 1$ or 2. We get two almost data-driven estimators of h: $\widehat{h}_{\widehat{m}_1}$ and $\widehat{h}_{\widehat{m}_2}$. Each penalty corresponds to a set of assumptions. Penalty pen_2^{th} corresponds to Assumption $\mathbf{A}_{\mathbf{mod}}^{(\mathbf{2})}$ so it works under both $\mathbf{A}_{\mathbf{mod}}^{(\mathbf{1})}$ and $\mathbf{A}_{\mathbf{mod}}^{(\mathbf{2})}$ (see remark 1). Penalty pen_2^{th} only works under Assumption $\mathbf{A}_{\mathbf{mod}}^{(\mathbf{1})}$, but is more computing-saving because $\overline{F}_0$ is estimated anyway, to compute the non adaptive estimators (see Δ_2^{th} in (11)).

Remark 2 Actually, any constant $B > 1$ could be allowed in the above penalties provided slight changes in the definition of Δ_2^{th}, but we fix $B > 3$ for simplicity's sake. This point is discussed more precisely in Section 5.5. Nevertheless, as B tends to 1, the constants C and C' involved in Theorem 1 tend to infinity.

3.3 Result

The following theorem states the adaptivity of $\widehat{h}_{\widehat{m}_i}$:

Theorem 1 *Let $i = 1$ or 2. Under Assumption $\mathbf{A}_{\mathbf{mod}}^{\mathbf{(i)}}$, $\mathbf{A}_{\mathbf{mod}}^{\mathbf{(3)}}$ and $\mathbf{A}_{\mathbf{frame}}$,*

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_i} - h\|^2_{\overline{F}_Y}\right] \leq C \inf_{m\in\mathcal{M}_n} \left\{\inf_{t\in S_m} \|t-h\|^2_{\overline{F}_Y} + pen_i^{th}(m)\right\} + \frac{C'}{n} \tag{13}$$

where C is a numerical constant and C' depend on $(K, \overline{F}_0, \|h\|_\infty)$

Comments:

1. We do not study explicitly the risk of $\widehat{h}_m$ for one model S_m but a particular case of (13) when $\mathcal{M}_n$ is restricted to $\{S_m\}$ provides the following inequality:
$$\mathbb{E}\left[\|\widehat{h}_m - h\|^2_{\overline{F}_Y}\right] \leq C" \left\{\inf_{t\in S_m} \|t-h\|^2_{\overline{F}_Y} + pen_i^{th}(m)\right\}$$
for $i = 1$ or 2.
2. Huber and MacGibbon (2004) prove that the minimax rate of convergence on the Hölder space $\mathcal{H}(\beta, L)$, for $\beta > 0$ and $L > 0$ is the classical rate $n^{-2\beta/(2\beta+1)}$. Besides, suppose that $h \in \mathcal{H}(\beta, L)$:
$$\inf_{t\in S_m} \|t-h\|_{\overline{F}_Y} \leq \inf_{t\in S_m} \|t-h\| \leq C(L,\beta) D_m^{-\beta}$$
Thus for a model of dimension $D_{m^*} = n^{1/(2\beta+1)}$:
$$\mathbb{E}\left[\|\widehat{h}_{m^*} - h\|^2_{\overline{F}_Y}\right] \leq C n^{-2\beta/(2\beta+1)}$$
So, for this choice D_{m^*}, $\widehat{h}_{m^*}$ is optimal in the minimax sense on the space $\mathcal{H}(\beta, L)$. Thus, the collection $\mathcal{M}_n$ contains an estimator with optimal rate, but the choice of $D_{m^*} = n^{1/(2\beta+1)}$ is not accessible as β is unknown.
3. The model selection procedure enables to choose automatically such a model, whithout estimating β. More precisely, Inequality (13) (called an oracle inequality) shows that the risk bound of $\widehat{h}_{\widehat{m}_i}$ has same order as the risk of the best estimator among the collection $\{\widehat{h}_m, m \in \mathcal{M}_n\}$. In particular, $\widehat{h}_{\widehat{m}_i}$ reaches the minimax rate of convergence $n^{2\beta/(2\beta+1)}$ over all Hölder classes $\mathcal{H}(\beta, L)$ for $\beta > 0$, $L > 0$.

4 Data-driven estimators

The estimators presented in this section are similar to the ones of Section 3, but the unknown quantities $\overline{F}_0$ and $\|h\|_{\infty,A}$ are replaced by estimators.

4.1 Estimator of $\overline{F}_0$

$\overline{F}_0$ is the lower bound of $\overline{F}_Y$ on $A = [0, 1]$, so $\overline{F}_0 = \overline{F}_Y(1)$. Thus a natural estimator of $\overline{F}_0$ would be the value of the empirical function in 1. For forcing the estimator of $\overline{F}_0$ to be lower bounded, we define:

$$\widehat{\overline{F}}_0 := \max(\alpha_n, \frac{1}{n}\sum_{i=1}^n 1_{Y_i\geq 1}), \qquad \text{where } \alpha_n = 1/\sqrt{n}$$

and denote by:

$$\Delta_3 = \left\{\frac{3}{4}\overline{F}_0 \leq \widehat{\overline{F}}_0 \leq \frac{5}{4}\overline{F}_0\right\}$$

4.2 Estimator of $\|h\|_{\infty,A}$

Let $\nu = \|h\|_{\infty,A}$. Let $D = E(n^\gamma)$ be a middle-sized model with $0 < \gamma < 1$, and $S_D = Vect(\varphi_1^D, \ldots, \varphi_1^D)$ the set of piecewise constant functions on $[0,1]$: $\varphi_j^D = \sqrt{D}1_{[\frac{j-1}{D},\frac{j}{D}[}$. Let $\widehat{h}_D := \arg\min_{t\in S_D}\gamma_n(t)$. As the basis functions (φ_j^D) have disjoint supports, the matrix $\widehat{G}_D$ of the scalar product $\langle .,.\rangle_n$ in the basis $(\varphi_1^D, \ldots, \varphi_D^D)$ is diagonal, with diagonal coefficients:

$$\left(\|\varphi_j^D\|_n^2\right)_{j=1,\ldots,D}$$

On the set $\varDelta_1$, these coefficients are positive, so the matrix $\widehat{G}_D$ is invertible and

$$\widehat{h}_D = \sum_{j=1,\ldots,D} \widehat{a}_j\varphi_j^D \text{ where } \widehat{a}_j = \frac{(1/n)\sum_{i=1}^n \delta_i\varphi_j^D(Y_i)}{\|\varphi_j^D\|_n^2}$$

Let us denote $\widehat{\nu}_n := \|\widehat{h}\|_\infty$, then:

$$\widehat{\nu}_n = \sqrt{D}\max_{j=1\ldots,D}\widehat{a}_j$$

Besides, we denote by h_D the $\|.\|_{\overline{F}_Y}$-projection of h on S_D, then

$$h_D = \sum_{j=1,\ldots,D} a_j\varphi_j \text{ where } a_j = \frac{\int_A \varphi_j^D(x)h(x)\overline{F}_Y(x)dx}{\|\varphi_j^D\|^2_{\overline{F}_Y(x)}} \tag{14}$$

Finally, we define the following set whose probability is close to 1 (see Proposition 63):

$$\varDelta_4 := \left\{\frac{3}{4}\nu \le \widehat{\nu}_n \le \frac{5}{4}\nu\right\}$$

4.3 Data-driven estimator

Let S_m be a model of the collection $\mathcal{M}_n$. We follow a procedure similar to the one described in Section 3.1, but now the set $\varDelta_2^{th}$ is replaced by

$$\varDelta_2 := \left\{\min(Sp(\widehat{G}_m)) \ge \frac{3}{5}\widehat{F}_0\right\}.$$

Now we have $\widehat{h}_m = \sum_{\lambda=1}^{D_m}\widehat{a}_\lambda^m\phi_\lambda$ with $\widehat{A}_m = (\widehat{a}_1^m, \ldots, \widehat{a}_{D_m}^m)^t$ given by $\widehat{A}_m = \widehat{G}_m^{-1}\widehat{V}_m$ on $\varDelta_2$, and 0 otherwise, where $\widehat{G}_m$ and $\widehat{V}_m$ are defined in (7).

Moreover, we take:

$$pen_1(m) = \frac{BK^2}{\widehat{F}_0}\frac{D_m}{n}, \quad pen_2(m) = B\widehat{\nu}_n\frac{D_m}{n}$$

with $B > 15/4$. Lastly we consider the estimators $\widehat{h}_{\widehat{m}_1}$ and $\widehat{h}_{\widehat{m}_2}$ where

$$\widehat{m}_i = \arg\min_{m\in\mathcal{M}_n}\gamma_n(\widehat{h}_m) + pen_i(m)$$

for $i = 1$ or 2.

4.4 Results

Now our estimators are completely data-driven when B is chosen, and we can generalize Theorem 1 as follows.

Theorem 2 *Let $i = 1$ or 2. Suppose that Assumption $\mathbf{A_{mod}^{(i)}}$, $\mathbf{A_{mod}^{(3)}}$ and $\mathbf{A_{frame}}$ hold, and that:*

$$\|h - h_D\|_\infty \leq \frac{\nu}{8} \tag{15}$$

where h_D is the projection of h in the histogram basis defined in (14)*, then:*

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_i} - h\|^2_{\overline{F}_Y}\right] \leq C \inf_{m\in\mathcal{M}_n} [\inf_{t\in S_m} \|t-h\|^2_{\overline{F}_Y} + pen_i^{th}(m)] + \frac{C'}{n}$$

where C is a numerical constant and C' depend on $(K, \overline{F}_0, \|h\|_\infty)$

Remark 3 1. If h is in the Hölder space $\mathcal{H}(\beta, L)$ for some $\beta \in]0,1[$, $L > 0$, then Assumption (15) is satisfied for n large enough. In fact, let $y \in A$:

$$\begin{aligned}
|h(y) - h_D(y)| &= \left| h(y) - \frac{D\int_{(j-1)/D}^{j/D} h(x)\overline{F}_Y(x)dx}{D\int_{(j-1)/D}^{j/D} \overline{F}_Y(x)dx} \right| \\
&= \frac{\left| D\int_{(j-1)/D}^{j/D} h(y)\overline{F}_Y(x)dx - D\int_{(j-1)/D}^{j/D} h(x)\overline{F}_Y(x)dx \right|}{D\int_{(j-1)/D}^{j/D} \overline{F}_Y(x)dx} \\
&\leq \frac{D\int_{(j-1)/D}^{j/D} |h(y) - h(x)|\overline{F}_Y(x)dx}{D\int_{(j-1)/D}^{j/D} \overline{F}_Y(x)dx} \leq \frac{L}{D^\beta}
\end{aligned}$$

The comments of Section 3.3 hold, so the adaptive estimators are minimax over Hölder spaces.

2. As notified in Remark 2, B could be choosen as any numerical constant, provided that it is greater than 1.

5 Proof of Theorem 1

The following Propositions are intermediate results to prove Theorem 1. Suppose that Assumption $\mathbf{A_{frame}}$ and $\mathbf{A_{mod}^{(3)}}$ hold.

Proposition 51 *Let $i = 1$ or 2. Suppose that Assumption $\mathbf{A_{mod}^{(i)}}$ holds, then*

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_i} - h\|^2_{\overline{F}_Y} 1_{\Delta_1}\right] \leq C \inf_{m\in\mathcal{M}_n} \left[\inf_{t\in S_m} \|t-h\|^2_{\overline{F}_Y} + pen_i^{th}(m)\right] + \frac{C'}{n} \tag{16}$$

where C is a numerical constant and C' depend on $(K, \overline{F}_0, \|h\|_\infty)$

Proposition 52 *For every model $S_m \in \mathcal{M}_n$:*

$$\|\widehat{h}_m - h\|_{\overline{F}_Y} \leq \frac{2K\sqrt{N_n}}{\overline{F}_0} + \|h\|_{\overline{F}_Y} \;\; a.e.$$

Proposition 53 *Suppose that Assumption $\mathbf{A_{mod}^{(1)}}$ or $\mathbf{A_{mod}^{(2)}}$ holds, then*

$$P(\Delta_1^c) \leq 2\exp(-C_2\overline{F}_0^2 \frac{n}{N_n})$$

where C_2 is a numerical constant.

5.1 Proof of Theorem 1

Under $\mathbf{A^{(i)}_{mod}}$, $N_n \leq n/(\ln n)^2$, so according to Proposition 52 and 53,

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_i} - h\|^2_{\overline{F}_Y} 1_{\Delta_1^c}\right] \leq \left(\frac{2K\sqrt{N_n}}{\overline{F}_0} + \|h\|_{\overline{F}_Y}\right)\exp[-C_2\overline{F}_0\frac{n}{N_n}] \leq Cn\exp[-C_2\overline{F}_0(\ln n)^2]$$
$$= Cn\left[\exp(-C_2\overline{F}_0\ln n)\right]^{\ln n} = Cn\left[n^{-C_2\overline{F}_0}\right]^{\ln n} = Cn^{1-C_2\overline{F}_0\ln n} \leq \frac{C'}{n} \quad (17)$$

And the result of Proposition 51 ends the proof of the Theorem. □

.

5.2 Proof of Proposition 51

Let i=1 or 2. To simplify the notations, we denote by $pen(m) = pen_i^{th}(m)$ and $\widehat{m} = \widehat{m}_i$. Let S_m be a model in the collection $\mathcal{M}_n$ and h_m be any function in S_m. On the set Δ_2^{th}, we have:

$$\gamma_n(\widehat{h}_{\widehat{m}}) + pen(\widehat{m}) \leq \gamma_n(h_m) + pen(m)$$

Thus,

$$\|\widehat{h}_{\widehat{m}}\|_n^2 - \|h_m\|_n^2 \leq pen(m) - pen(\widehat{m}) + \frac{2}{n}\sum_{i=1}^n \delta_i(\widehat{h}_{\widehat{m}} - h_m)(Y_i)$$

Besides, $\|\widehat{h}_{\widehat{m}} - h_m\|_n^2 = \|\widehat{h}_{\widehat{m}}\|_n^2 + \|h_m\|_n^2 - 2\langle\widehat{h}_{\widehat{m}}, h_m\rangle_n$, so:

$$\|\widehat{h}_{\widehat{m}} - h_m\|_n^2 \leq pen(m) - pen(\widehat{m}) + \frac{2}{n}\sum_{i=1}^n \delta_i(\widehat{h}_{\widehat{m}} - h_m)(Y_i) - 2\langle\widehat{h}_{\widehat{m}}, h_m\rangle_n + 2\|h_m\|_n^2$$
$$= pen(m) - pen(\widehat{m}) - 2\langle\widehat{h}_{\widehat{m}} - h_m, h_m\rangle_n + \frac{2}{n}\sum_{i=1}^n \delta_i(\widehat{h}_{\widehat{m}} - h_m)(Y_i)$$
$$= pen(m) - pen(\widehat{m}) - 2\langle\widehat{h}_{\widehat{m}} - h_m, h_m - h\rangle_n + 2\nu_n(\widehat{h}_{\widehat{m}} - h_m)$$

where we denote:

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n \delta_i t(Y_i) - \langle t, h\rangle_n$$

Let us denote $S_m + S_{m'} = \{t + t', t \in S_m, t' \in S_{m'}\}$, then by noting that $2ab \leq 2a^2 + (1/2)b^2$ for every a, b, and with Cauchy-Schwartz Inequality:

$$\|\widehat{h}_{\widehat{m}} - h_m\|_n^2 \leq pen(m) - pen(\widehat{m}) - 2\langle\widehat{h}_{\widehat{m}} - h_m, h_m - h\rangle_n + 2\|\widehat{h}_{\widehat{m}} - h_m\|_{\overline{F}_Y} \sup_{t\in S_m + S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} \nu_n(t)$$
$$\leq pen(m) - pen(\widehat{m}) + 2\|\widehat{h}_{\widehat{m}} - h_m\|_n\|h_m - h\|_n + \frac{1}{2}\|\widehat{h}_{\widehat{m}} - h_m\|^2_{\overline{F}_Y}$$
$$+2\sup_{t\in S_m + S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} (\nu_n(t))^2$$

Since $2ab \leq (1/4)a^2 + 4b^2$, for every $p(m, m')$ function of (m, m'):

$$\|\widehat{h}_{\widehat{m}} - h_m\|_n^2 \leq pen(m) - pen(\widehat{m}) + 2p(m, \widehat{m}) + 2\|\widehat{h}_{\widehat{m}} - h_m\|_n\|h_m - h\|_n + \frac{1}{2}\|\widehat{h}_{\widehat{m}} - h_m)\|^2_{\overline{F}_Y}$$
$$+2\sup_{t\in S_m + S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} [(\nu_n(t))^2 - p(m, \widehat{m})]$$
$$\leq pen(m) - pen(\widehat{m}) + 2p(m, \widehat{m}) + \frac{1}{4}\|\widehat{h}_{\widehat{m}} - h_m\|_n^2 + 4\|h - h_m\|_n^2$$
$$+\frac{1}{2}\|\widehat{h}_{\widehat{m}} - h_m\|^2_{\overline{F}_Y} + 2\sup_{t\in S_m + S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} [(\nu_n(t))^2 - p(m, \widehat{m})]$$

Thus:

$$\frac{3}{4}\|\widehat{h}_{\widehat{m}} - h_m\|_n^2 \le pen(m) - pen(\widehat{m}) + 2p(m,\widehat{m}) + \frac{1}{2}\|\widehat{h}_{\widehat{m}} - h_m\|^2_{\overline{F}_Y} + 4\|h - h_m\|^2_{\overline{F}_Y}$$
$$+2 \sup_{t\in S_m+S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} [(\nu_n(t))^2 - p(m,\widehat{m})]$$

On the set $\varDelta_1 \cap \varDelta_2^{th} = \varDelta_1$ (cf Section 3.1), $\|\widehat{h}_{\widehat{m}} - h_m\|_n \ge \frac{3}{4}\|\widehat{h}_{\widehat{m}} - h_m\|_{\overline{F}_Y}$, hence

$$((\frac{3}{4})^2 - \frac{1}{2})\|\widehat{h}_{\widehat{m}} - h_m\|^2_{\overline{F}_Y} \le 4\|h - h_m\|_n^2 + pen(m) - pen(\widehat{m}) + 2p(m,\widehat{m})$$
$$+2 \sup_{t\in S_m+S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} [(\nu_n(t))^2 - p(m,\widehat{m})]$$

Moreover, we note that $\|\widehat{h}_{\widehat{m}} - h_m\|^2_{\overline{F}_Y} \ge \frac{1}{2}\|\widehat{h}_{\widehat{m}} - h\|^2_{\overline{F}_Y} - \|h - h_m\|^2_{\overline{F}_Y}$, and $\mathbb{E}\left[\|h-h_m\|_n^2\right] = \|h - h_m\|^2_{\overline{F}_Y}$ so:

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}} - h\|^2_{\overline{F}_Y} 1_{\varDelta_1}\right] \le C_1 \Big\{\|h - h_m\|^2_{\overline{F}_Y} + \mathbb{E}[pen(m) - pen(\widehat{m}) + 2p(m,\widehat{m})] +$$
$$\sup_{t\in S_m+S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} [(\nu_n(t))^2 - p(m,\widehat{m})]\Big\} \tag{18}$$

where C_1 is a numerical constant. Besides, $\nu_n(t)$ is a centered process since

$$\mathbb{E}[\delta_i t(Y_i)] = \int_A t(x)h(x)\overline{F}_Y(x)dx = \mathbb{E}[\langle t, h\rangle_n]$$

(see (6)). Therefore, we insert the mean term $\int_A t(x)h(x)\overline{F}_Y(x)dx$ to obtain the sum of two variance-type terms. More precisely, we define:

$$\nu_{n,1}(t) = \frac{1}{n}\sum_{i=1}^n \delta_i t(Y_i) - \int_A t(x)h(x)\overline{F}_Y(x)dx$$
$$\nu_{n,2}(t) = \frac{1}{n}\sum_{i=1}^n \int_A t(x)h(x)1_{Y_i\ge x}dx - \int_A t(x)h(x)\overline{F}_Y(x)dx$$

Then, since $(a+b)^2 \le \frac{3}{2}a^2 + 3b^2$,

$$\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} ((\nu_n(t))^2 - p(m,\widehat{m}))\right]$$
$$\le \frac{3}{2}\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} ((\nu_{n,1}(t))^2 - \frac{2}{3}p(m,\widehat{m}))_+\right] + 3\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}}, \|t\|_{\overline{F}_Y}=1} (\nu_{n,2}(t))^2\right] \tag{19}$$

Moreover, the two terms above are upper-bounded as follows.

lemma 51 *Under the Assumptions of Theorem* 1

$$\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}_i}, \|t\|_{\overline{F}_Y}=1} (\nu_{n,2}(t))^2\right] \le \frac{1}{\overline{F}_0 n}\|h\|^2_{\overline{F}_Y}$$

lemma 52 *1) Let*

$$p_1(m,m') = \frac{BK^2}{2\overline{F}_0}\frac{D_m + D_{m'}}{n} \tag{20}$$

then under Assumptions $\mathbf{A}^{(1)}_{\mathbf{mod}}$, $\mathbf{A}^{(3)}_{\mathbf{mod}}$ *and* $\mathbf{A}_{\mathbf{frame}}$, *we have:*

$$\mathbb{E}\left[\left(\sup_{t\in S_m+S_{\widehat{m}_1}, \|t\|_{\overline{F}_Y}=1} (\nu_{n,1}(t))^2 - \frac{2}{3}p_1(m,\widehat{m}_1)\right)_+\right] \le \frac{C'}{n}$$

2) Let

$$p_2(m,m') = \frac{BK^2\|h\|_{\infty,A}}{2}\frac{D_m + D_{m'}}{n} \tag{21}$$

then under Assumptions $\mathbf{A^{(1)}_{mod}}$*,* $\mathbf{A^{(3)}_{mod}}$ *and* $\mathbf{A_{frame}}$*, we have:*

$$\mathbb{E}\left[\left(\sup_{t\in S_m+S_{\widehat{m}_2},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2 - \frac{2}{3}p_2(m,\widehat{m}_2)\right)_+\right] \leq \frac{C'}{n}$$

for some constant C' *depending on* $(K, \|h\|_{\infty,A}, \overline{F}_0)$.

Finally, for the estimator $\widehat{h}_{\widehat{m}_1}$:

$$pen(m) - pen(\widehat{m}_1) + 2p_1(m,\widehat{m}_1) = \frac{B\theta K^2}{\overline{F}_0}\frac{D_m}{n} \tag{22}$$

and for the estimator $\widehat{h}_{\widehat{m}_2}$

$$pen(m) - pen(\widehat{m}_1) + 2p_1(m,\widehat{m}_1) = B\theta K^2\|h\|_{\infty,A}\frac{D_m}{n} \tag{23}$$

Finally inequalities (18), (19), (22) or (23), and the results of Lemmas 51 and 52 provides inequality (16), which ends the proof of 51. □

Proof of Lemma 51

Let $(\phi_1^n,\dots,\phi_{N_n}^n)$ be an $\|.\|$-orthonormal basis of the global space S_n, and note that $\{t, \|t\|^2_{\overline{F}_Y}\leq 1\} \subset \{t, \|t\|^2 \leq \overline{F}_0^{-1}\}$. Then:

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}_i},\|t\|_{\overline{F}_Y}=1}(\nu_{n,2}(t))^2\right] &\leq \mathbb{E}\left[\sup_{t\in S_n,\|t\|_{\overline{F}_Y}=1}(\nu_{n,2}(t))^2\right]\\
&\leq \mathbb{E}\left[\sup_{t\in S_n,\|t\|^2\leq 1/\overline{F}_0}(\nu_{n,2}(t))^2\right]\\
&= \mathbb{E}\left[\sup_{\sum_{\lambda=1}^{N_n}a_\lambda^2\leq 1/\overline{F}_0}\left(\sum_{\lambda=1}^{N_n}a_\lambda(\langle\phi_\lambda^n,h\rangle_n - \langle\phi_\lambda^n,h\rangle_{\overline{F}_Y})\right)^2\right]
\end{aligned}$$

With Cauchy-Schwartz Inequality, we obtain

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}_i},\|t\|_{\overline{F}_Y}=1}(\nu_{n,2}(t))^2\right] &\leq \frac{1}{\overline{F}_0}\sum_{\lambda=1}^{N_n}\mathbb{E}\left[(\langle\phi_\lambda^n,h\rangle_n - \langle\phi_\lambda^n,h\rangle_{\overline{F}_Y})^2\right] = \frac{1}{\overline{F}_0}\sum_{\lambda=1}^{N_n}\frac{1}{n}Var\left[\int_A h(x)\phi_\lambda^n(x)1_{Y_1\geq x}dx\right]\\
&\leq \frac{1}{\overline{F}_0 n}\mathbb{E}\left[\sum_{\lambda=1}^{N_n}\langle\phi_\lambda^n, h(.)1_{Y_1\geq .}\rangle^2\right] = \frac{1}{\overline{F}_0 n}\mathbb{E}\left[\|(h(.)1_{Y_1\geq .})_{S_n}\|^2\right]
\end{aligned}$$

where $(h(.)1_{Y_1\geq .})_{S_n}$ denotes the L^2-orthogonal projection of $h(.)1_{Y_1\geq .}$ on S_n. Thus:

$$\mathbb{E}\left[\sup_{t\in S_m+S_{\widehat{m}_i},\|t\|_{\overline{F}_Y}=1}(\nu_{n,2}(t))^2\right] \leq \frac{1}{\overline{F}_0 n}\mathbb{E}\left[\|h(.)1_{Y_1\geq .}\|^2\right] = \frac{1}{\overline{F}_0 n}\|h\|^2_{\overline{F}_Y} \quad \square$$

Proof of Lemma 52

For i=1 or 2, we have:

$$\mathbb{E}\left[\left(\sup_{t\in S_m+S_{\widehat{m}_i},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2 - \frac{2}{3}p_i(m,\widehat{m}_i)\right)_+\right] \leq \sum_{m'\in J_n}\mathbb{E}\left[\left(\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2 - \frac{2}{3}p_i(m,m')\right)_+\right]$$

Besides, for every models $S_m, S_{m'}$, we upper bound the term $\mathbb{E}[(\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2-p_i(m,m'))_+]$ with Talagrand Inequality as recalled in Theorem 4.

Let us compute first the term $\mathbb{H}$. Under Assumption $\mathbf{A^{(1)}_{mod}}$, $S_m \subset S_{m'}$ or $S_{m'} \subset S_m$. Thus, if we denote by $D_{m+m'}$ the dimension of $S_m + S'_m$, we have $D_{m+m'} = \max(D_m, D_{m'}) \leq D_m + D_{m'}$. Moreover, we denote by $(\phi_1^{m+m'}, \ldots, \phi_{D_{m+m'}}^{m+m'})$ the orthonormal basis of $S_m + S_{m'}$ with $\phi_\lambda^{m+m'} = \phi_\lambda^m$ if $S_m \subset S_{m'}$, and $\phi_\lambda^{m'}$ if $S_{m'} \subset S_m$.

1. Suppose that the assumptions of 1. in Lemma 52 hold. Similarly to the upper bound of $\mathbb{E}[\sup_{t\in S_m+S_{\widehat{m}},\|t\|_{\overline{F}_Y}=1}(\nu_{n,2}(t))^2]$, we obtain

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2\right] &\leq \mathbb{E}\left[\sup_{t\in S_m+S_{m'},\|t\|\leq 1/\overline{F}_0}(\nu_{n,1}(t))^2)_+\right]\\ &\leq \frac{1}{\overline{F}_0}\sum_{\lambda=1}^{D_{m+m'}}\frac{1}{n}Var\left[\delta_1\phi_\lambda^{m+m'}(Y_1)\right]\\ &\leq \frac{1}{\overline{F}_0}\sum_{\lambda=1}^{D_{m+m'}}\frac{1}{n}\mathbb{E}\left[(\phi_\lambda^{m+m'})^2(Y_1)\right] \qquad (24)\end{aligned}$$

Besides, under Assumption $\mathbf{A^{(1)}_{mod}}$:

$$\left\|\sum_{\lambda=1}^{D_{m+m'}}(\phi_\lambda^{m+m'})^2\right\|_\infty \leq K^2 D_{m+m'}, \qquad (25)$$

hence:

$$\mathbb{E}\left[\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2\right] \leq \frac{1}{n\overline{F}_0}\left\|\sum_{\lambda=1}^{D_{m+m'}}(\phi_\lambda^{m+m'})^2\right\|_\infty \leq \frac{K^2(D_m+D_{m'})}{\overline{F}_0 n} := \mathbb{H}^2$$

Since $B > 3$, according to the definition (20) of $p_1(m,m')$, we have $(2/3)p(m,m') = \theta\mathbb{H}^2$ for some $\theta > 1$. Besides,

$$\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}\|\delta_1 t(Y_1)\|_\infty \leq \sup_{t\in S_m+S_{m'},\|t\|^2\leq 1/\overline{F}_0}\|t\|_\infty = \sup_{\sum_{\lambda=1}^{D_{m+m'}}a_\lambda^2\leq 1/\overline{F}_0}\left\|\sum_{\lambda=1}^{D_{m+m'}}a_\lambda\phi_\lambda^{m+m'}\right\|_\infty$$

With Cauchy-Schwartz Inequality, and inequality (25), we obtain

$$\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}\|\delta_1 t(Y_1)\|_\infty \leq \frac{1}{\sqrt{\overline{F}_0}}\left\|\sum_{\lambda=1}^{D_{m+m'}}(\phi_\lambda^{m+m'})^2\right\|_\infty \leq \frac{K}{\sqrt{\overline{F}_0}}\sqrt{D_m+D_{m'}} := b$$

And according to computings (6),

$$\begin{aligned}\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}Var(\delta_1 t(Y_1)) &\leq \sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}\mathbb{E}[\delta_1 t^2(Y_1)]\\ &= \sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}\int_A t^2(x)h(x)\overline{F}_Y(x)dx \leq \|h\|_{\infty,A} := v\end{aligned}$$

Then, with Talagrand Inequality, we obtain

$$\begin{aligned}&\mathbb{E}\left[\left(\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2 - \frac{2}{3}p(m,m')\right)_+\right]\\ &\leq \overline{C}\frac{\|h\|_{\infty,A}}{n}\exp\left(-\kappa\frac{K^2(D_m+D_{m'})}{\overline{F}_0\|h\|_{\infty,A}}\right) + \overline{C}'\frac{K^2(D_m+D_{m'})}{\overline{F}_0^2 n^2}\exp(-\kappa'\sqrt{n})\end{aligned}$$

Thus, with Assumption $\mathbf{A^{(3)}_{mod}}$, we have

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2-\frac{2}{3}p(m,m')\right)_+\right]\leq\frac{C'}{n}$$

which conclude the proof of 1. in Lemma 52.

2. Suppose that the assumptions of 2. in Lemma 52. Let $(\psi_1,\dots,\psi_{D_{m+m'}})$ be an orthonormal basis of $S_m+S_{m'}$ for the norm $\|.\|_{\overline{F}_Y}$. Then, similarly to (24):

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2\right]&\leq\frac{1}{n}\sum_{\lambda=1}^{D_{m+m'}}Var(\delta_1\psi_\lambda^2(X_1))\leq\frac{1}{n}\sum_{\lambda=1}^{D_{m+m'}}\mathbb{E}[\delta_1\psi_\lambda^2(X_1)]\\&=\frac{1}{n}\sum_{\lambda=1}^{D_{m+m'}}\int_A\psi_\lambda^2(x)h(x)\overline{F}_Y(x)dx\\&\leq\|h\|_{\infty,A}\frac{D_m+D_{m'}}{n}:=\mathbb{H}^2\end{aligned}$$

Besides, according to Assumption $\mathbf{A^{(2)}_{mod}}$:

$$\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}\|\delta_1t(Y_1)\|_\infty\leq\sup_{t\in S_m+S_{m'},\|t\|^2\leq1/\overline{F}_0}\|t\|_\infty\leq\frac{K}{\sqrt{\overline{F}_0}}\sqrt{N_n}:=b$$

and the end of the proof is similar to the 1). $\square$

5.3 Proof of Proposition 52

Let $m\leq N_n$:

$$\begin{aligned}\|\widehat{h}_m-h\|_{\overline{F}_Y}&\leq\|\widehat{h}_m\|+\|h\|_{\overline{F}_Y}\\&=\|\widehat{A}_m\|+\|h\|_{\overline{F}_Y}\\&\leq\max\left(Sp(\widehat{G}_m^{-1})\right)\|\widehat{V}_m\|+\|h\|_{\overline{F}_Y}\\&=\left[\min\left(Sp(\widehat{G}_m)\right)\right]^{-1}\|\widehat{V}_m\|+\|h\|_{\overline{F}_Y}\end{aligned}$$

According to inequality (10):

$$\begin{aligned}\|\widehat{h}_m-h\|_{\overline{F}_Y}&\leq\frac{4}{3\overline{F}_0}[\sum_{\lambda=1}^{D_m}(\frac{1}{n}\sum_{i=1}^n\phi_\lambda^m(Y_i)\delta_i)^2]^{1/2}+\|h\|_{\overline{F}_Y}\\&\leq\frac{4}{3\overline{F}_0}[\sum_{\lambda=1}^{D_m}\frac{1}{n}\sum_{i=1}^n(\phi_\lambda^m)^2(Y_i)]^{1/2}+\|h\|_{\overline{F}_Y}\\&\leq\frac{4}{3\overline{F}_0}\left\|\sum_{\lambda=1}^{D_m}(\phi_\lambda^m)^2\right\|_\infty^{1/2}+\|h\|_{\overline{F}_Y}\\&\leq\frac{4K\sqrt{N_n}}{3\overline{F}_0}+\|h\|_{\overline{F}_Y}\quad\square\end{aligned}$$

5.4 Proof of Proposition 53

The proof of Proposition 53 is inspired from Baraud (2002).

$$\varDelta_1^c = \left\{ |\|t\|_n^2 - \|t\|_{\overline{F}_Y}^2| > \frac{1}{4}\|t\|_{\overline{F}_Y}^2, \forall t \in S_n \right\} = \left\{ \sup_{t\in S_n, \|t\|_{\overline{F}_Y}=1} \eta_n(t^2) > \frac{1}{4} \right\}$$

where $\eta_n(t) = \frac{1}{n}\sum_{i=1}^n (\int_A t(x)1_{Y_i\geq x}dx - \int_A t(x)\overline{F}_Y(x)dx)$. Let $(\psi_1, \ldots, \psi_{N_n})$ be an orthonormal base of the global space S_n for the norm $\|.\|_{\overline{F}_Y}$, then

$$\begin{aligned}
\varDelta_1^c &= \left\{ \sup_{\sum a_\lambda^2=1} \sum_{\lambda,\lambda'} a_\lambda a_{\lambda'} \left[\frac{1}{n}\sum_{i=1}^n (\int_A \psi_\lambda(x)\psi_{\lambda'}(x)1_{Y_i\geq x}dx - \int_A \psi_\lambda(x)\psi_{\lambda'}(x)\overline{F}_Y(x)dx) \right] > \frac{1}{4} \right\} \\
&:= \left\{ \sup_{\sum a_\lambda^2=1} \sum_{\lambda,\lambda'} a_\lambda a_{\lambda'} S_{\lambda,\lambda'} > \frac{1}{4} \right\}
\end{aligned}$$

On the one hand, let λ, λ' be fixed, then:

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\left(\int_A \psi_\lambda\psi_{\lambda'}1_{Y_i\geq x}dx \right)^2 \right] = \mathbb{E}\left[\left(\int_A \psi_\lambda\psi_{\lambda'}1_{Y_1\geq x}dx \right)^2 \right] := v_{\lambda,\lambda'}$$

and for every $l \geq 2$:

$$\begin{aligned}
\mathbb{E}\left[\left(\int_A \psi_\lambda(x)\psi_{\lambda'}(x)1_{Y_1\geq x}dx \right)_+^l \right] &\leq \mathbb{E}\left[\left(\int_A \psi_\lambda(x)\psi_{\lambda'}(x)1_{Y_1\geq x}dx \right)^2 \left(\int_A |\psi_\lambda(x)\psi_{\lambda'}(x)|dx \right)^{l-2} \right] \\
&\leq v_{\lambda,\lambda'} \left(\int_A \psi_\lambda^2(x)dx \int_A \psi_{\lambda'}^2(x)dx \right)^{l/2-1} \\
&\leq v_{\lambda,\lambda'} \left(\frac{1}{\overline{F}_0} \right)^{l-2} \left(\int_A \psi_\lambda^2(x)\overline{F}_Y(x)dx \int_A \psi_{\lambda'}^2(x)\overline{F}_Y(x)dx \right)^{l/2-1} \\
&= v_{\lambda,\lambda'} \left(\frac{1}{\overline{F}_0} \right)^{l-2} := v_{\lambda,\lambda'} c_{\lambda,\lambda'}^{l-2}
\end{aligned}$$

Thus, Bernstein Inequality recalled in Theorem 3 provides the following upper bound:

$$P\left[|S_{\lambda,\lambda'}| \geq \sqrt{2v_{\lambda,\lambda'}x} + c_{\lambda,\lambda'}x \right] \leq 2\exp(-nx)$$

On the other hand,

$$\begin{aligned}
&\left\{ |S_{\lambda,\lambda'}| < \sqrt{2v_{\lambda,\lambda'}x} + c_{\lambda,\lambda'}x, \forall \lambda, \lambda' = 1, \ldots N_n \right\} \\
&\subset \left\{ \sum_{\lambda,\lambda'=1}^{N_n} |a_\lambda||S_{\lambda,\lambda'}||a_{\lambda'}| < \sqrt{2x} \sum_{\lambda,\lambda'=1}^{N_n} |a_\lambda|\sqrt{v_{\lambda,\lambda'}}|a_{\lambda'}| + x \sum_{\lambda,\lambda'=1}^{N_n} |a_\lambda|c_{\lambda,\lambda'}|a_{\lambda'}|, \forall (a_\lambda)_{\lambda=1,\ldots N_n} \right\} \\
&\subset \left\{ \sup_{\sum a_\lambda^2=1} \sum_{\lambda,\lambda'=1}^{N_n} |a_\lambda||S_{\lambda,\lambda'}||a_{\lambda'}| < \sqrt{2x} \sup_{\sum a_\lambda^2=1} \sum_{\lambda,\lambda'=1}^{N_n} |a_\lambda|\sqrt{v_{\lambda,\lambda'}}|a_{\lambda'}| + x \sup_{\sum a_\lambda^2=1} \sum_{\lambda,\lambda'=1}^{N_n} |a_\lambda|c_{\lambda,\lambda'}|a_{\lambda'}| \right\} \\
&= \left\{ \sup_{t\in S_n, \|t\|_{\overline{F}_Y}\leq 1} |\eta_n(t^2)| \leq \sqrt{2x}\rho(V) + x\rho(C) \right\}
\end{aligned}$$

where $\rho(M)$ denotes the maximum of the spectrum of M, and V and C denote the following matrix:

$$V := \left(\sqrt{v_{\lambda,\lambda'}}\right)_{\lambda,\lambda'=1,\ldots,N_n}, \quad C = (c_{\lambda,\lambda'})_{\lambda,\lambda'=1,\ldots,N_n}$$

Thus, for every $x \geq 0$,

$$P\left[\sup_{t\in S_n, \|t\|_{\overline{F}_Y}\leq 1} |\eta_n(t^2)| > \sqrt{2x}\rho(V) + x\rho(C)\right] \leq \sum_{\lambda,\lambda'} P\left[|S_{\lambda,\lambda'}| > \sqrt{2v_{\lambda,\lambda'}x} + c_{\lambda,\lambda'}x\right] \leq 2N_n^2 \exp(-nx)$$

To upper bound the term $P[\Delta_1^c]$, we choose x such that $\sqrt{2x}\rho(V) \leq \frac{1}{8}$ and $x\rho(C) \leq \frac{1}{8}$. Let $L(\psi) = \max(\rho(C), 16\rho(V)^2)$ then,

$$P[\Delta_1^c] \leq 2\exp\left(-\frac{n}{8L(\psi)}\right)$$

Now we have to upper bound $L(\psi)$. Applying two times Cauchy-Schwartz Inequality,we obtain

$$\begin{aligned}
(\rho(V))^2 &= \sup_{\sum a_\lambda^2=1}\left[\sum_{\lambda=1}^{N_n}|a_\lambda|\left(\sum_{\lambda'=1}^{N_n}|a_{\lambda'}|\sqrt{v_{\lambda,\lambda'}}\right)\right]^2 \leq \sup_{\sum a_\lambda^2=1}\left(\sum_{\lambda=1}^{N_n}a_\lambda^2\right)\left(\sum_{\lambda=1}^{N_n}\left[\sum_{\lambda'=1}^{N_n}||a_{\lambda'}|\sqrt{v_{\lambda,\lambda'}}\right]^2\right)\\
&= \sup_{\sum a_\lambda^2=1}\sum_{\lambda=1}^{N_n}\left(\sum_{\lambda'=1}^{N_n}|a_{\lambda'}|\sqrt{v_{\lambda,\lambda'}}\right)^2 \leq \sum_{\lambda=1}^{N_n}\left(\sum_{\lambda'=1}^{N_n}v_{\lambda,\lambda'}\right)
\end{aligned}$$

We replace $v_{\lambda,\lambda'}$ by its expression.

$$\begin{aligned}
(\rho(V))^2 &\leq \sum_{\lambda=1}^{N_n}\mathbb{E}\left[\sum_{\lambda'=1}^{N_n}\langle\psi_{\lambda'}, \psi_\lambda 1_{Y_1\geq .}\rangle^2\right]\\
&\leq \frac{1}{\overline{F}_0}\sum_{\lambda=1}^{N_n}\mathbb{E}\left[\sum_{\lambda'=1}^{N_n}\langle\psi_{\lambda'}, \psi_\lambda 1_{Y_1\geq .}\rangle^2_{\overline{F}_Y}\right]
\end{aligned}$$

Besides, $\sqrt{\sum_{\lambda'=1}^{N_n}\langle\psi_{\lambda'}, \psi_\lambda 1_{Y_1\geq x}\rangle^2_{\overline{F}_Y}}$ is equal to the norm of the $\|.\|_{\overline{F}_Y}$-projection of $\psi_\lambda 1_{Y_1\geq .}$ on S_n, so:

$$\sum_{\lambda'=1}^{N_n}\langle\psi_{\lambda'}, \psi_\lambda 1_{Y_1\geq x}\rangle^2_{\overline{F}_Y} \leq \|\psi_\lambda 1_{Y_1\geq .}\|^2_{\overline{F}_Y} \leq \|\psi_\lambda\|^2_{\overline{F}_Y} = 1$$

Hence $(\rho(V))^2 \leq \frac{N_n}{\overline{F}_0^2}$.

Besides,

$$\begin{aligned}
\rho(C) &= \frac{1}{\overline{F}_0}\sup_{\sum a_\lambda^2=1}\left(\sum_{\lambda,\lambda'=1}^{N_n}|a_\lambda||a_{\lambda'}|\right)\\
&\leq \frac{1}{\overline{F}_0}\sup_{\sum a_\lambda^2=1}\sqrt{\sum_{\lambda=1}^{N_n}a_\lambda^2}\sqrt{\sum_{\lambda=1}^{N_n}(\sum_{\lambda'=1}^{N_n}|a_{\lambda'}|)^2}\\
&\leq \frac{1}{\overline{F}_0}\sup_{\sum a_\lambda^2=1}\sqrt{N_n}\sum_{\lambda'=1}^{N_n}|a_{\lambda'}| \leq \frac{N_n}{\overline{F}_0}
\end{aligned}$$

Finally $L(\psi) \leq max(\frac{N_n}{\overline{F}_0}, 16\frac{N_n}{\overline{F}_0^2}) = 16\frac{N_n}{\overline{F}_0^2}$ and

$$P[\Delta_1^c] \leq \exp\left(-C_2\overline{F}_0^2\frac{n}{N_n}\right) \quad \square$$

5.5 Comment about the constant in the penalty

This Section ensues from Remark 2. Suppose that Δ_2^{th} is replaced by the set:

$$\left\{|\frac{\|t\|_n^2}{\|t\|^2_{\overline{F}_Y}} - 1| \leq \alpha, \forall t \in S_n\right\}$$

and the inequalities of the kind $2ab \leq 2a^2 + (1/2)b^2$ by $2ab \leq (1/\beta)a^2 + \beta b^2$ in the proofs above. Then if α, β are chosen small enough, Theorem 1 still holds for a constant $B > 1$ in the penalty.

6 Proof of Theorem 2

The following Propositions are intermediate results to prove Theorem 2.

Proposition 61 *Suppose that Assumptions* $\mathbf{A}^{(3)}_{\mathbf{mod}}$ *and* $\mathbf{A}_{\mathbf{frame}}$ *hold.*

1. *If Assumption* $\mathbf{A}^{(1)}_{\mathbf{mod}}$ *holds, then:*
$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_1} - h\|^2_{\overline{F}_Y} 1_{\Delta_1\cap\Delta_2\cap\Delta_3}\right] \leq C \inf_{m\in\mathcal{M}_n}\left[\inf_{t\in S_m}\|t-h\|^2_{\overline{F}_Y} + pen^{th}_1(m)\right] + \frac{C'}{n}$$
2. *If Assumption* $\mathbf{A}^{(2)}_{\mathbf{mod}}$ *holds, then:*
$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_2} - h\|^2_{\overline{F}_Y} 1_{\Delta_1\cap\Delta_2\cap\Delta_3\cap\Delta_4}\right] \leq C \inf_{m\in\mathcal{M}_n}\left[\inf_{t\in S_m}\|t-h\|^2_{\overline{F}_Y} + pen^{th}_2(m)\right] + \frac{C'}{n}$$
where C *is a numerical constant and* C' *depends on* $(K, \overline{F}_0, \|h\|_\infty)$

Proposition 62 *Under Assumption* $\mathbf{A}_{\mathbf{frame}}$*:*

1. *For every* n *such that* $\alpha_n \leq \overline{F}_0/2$*:*
$$P[\Delta_3^c] \leq 2\exp(-C_1 n\overline{F}_0)$$
for some numerical constant C_1*.*
2. $$\Delta_2^c \cap \Delta_3 \subset \Delta_1^c \cap \Delta_3$$

Proposition 63 *Under Assumptions* (15) *and* $\mathbf{A}_{\mathbf{frame}}$*, we have:*

$$P[\Delta_4^c \cap \Delta_1] \leq 4D\exp\left(-C\frac{n}{D}\right) \tag{26}$$

where C *depends on* $(\nu, \overline{F}_0, \|h\|_\infty)$*.*

6.1 Proof of Theorem 2

Similarly to the proof of Proposition 52, for every model m,

$$\|\widehat{h}_m - h\|_{\overline{F}_Y} \leq \frac{2K\sqrt{N_n}}{\widehat{F}_0} + \|h\|_{\overline{F}_Y} \leq \frac{2K\sqrt{N_n}}{\alpha_n} + \|h\|_{\overline{F}_Y} = 2Kn + \|h\|_{\overline{F}_Y}$$

since $\widehat{F}_0 \geq \alpha_n$ and $N_n \leq n$.

- Proof for $i = 1$:
$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_1} - h\|^2_{\overline{F}_Y} 1_{(\Delta_1\cap\Delta_2\cap\Delta_3)^c}\right] \leq \left(2Kn + \|h\|_{\overline{F}_Y}\right)^2 P\left[(\Delta_1\cap\Delta_2\cap\Delta_3)^c\right]$$
and
$$\begin{aligned} P\left[(\Delta_1\cap\Delta_2\cap\Delta_3)^c\right] &= P\left[\Delta_1^c \cup \Delta_2^c \cup \Delta_3^c\right] \\ &= P\left[(\Delta_1^c \cup \Delta_2^c \cup \Delta_3^c)\cap\Delta_3\right] + P\left[(\Delta_1^c \cup \Delta_2^c \cup \Delta_3^c)\cap\Delta_3^c\right] \\ &\leq P\left[(\Delta_1^c\cap\Delta_3)\cup(\Delta_2^c\cap\Delta_3)\right] + P\left[\Delta_3^c\right] \end{aligned}$$
According to 2) of Proposition 62, we have
$$P\left[(\Delta_1\cap\Delta_2\cap\Delta_3)^c\right] \leq P\left[\Delta_1^c\cap\Delta_3\right] + P\left[\Delta_3^c\right] \leq P\left[\Delta_1^c\right] + P\left[\Delta_3^c\right]$$
Thus Proposition 53 with computings in (17), and Proposition 62 lead to
$$P\left[(\Delta_1\cap\Delta_2\cap\Delta_3)^c\right] \leq \exp\left(-C_1\overline{F}_0\frac{n}{\ln n}\right) + 2\exp(-C_2\overline{F}_0 n)$$
Hence $\mathbb{E}\left[\|\widehat{h}_{m_1} - h\|^2_{\overline{F}_Y} 1_{(\Delta_1\cap\Delta_2\cap\Delta_3)^c}\right] \leq \frac{C'}{n}$ and 1. of Proposition 61 ends the proof for $i = 1$.

- Proof for $i=2$:

$$
\begin{aligned}
P\left[(\Delta_1\cap\Delta_2\cap\Delta_3\cap\Delta_4)^c\right] &= P\left[(\Delta_1^c\cup\Delta_2^c\cup\Delta_3^c\cup\Delta_4^c)\cap\Delta_1\right]+P\left[\Delta_1^c\right]\\
&\leq P\left[\Delta_1^c\cup\Delta_2^c\cup\Delta_3^c\right]+P\left[\Delta_4^c\cap\Delta_1\right]+P\left[\Delta_1^c\right]
\end{aligned}
$$

According to inequality (27),

$$
P\left[(\Delta_1\cap\Delta_2\cap\Delta_3\cap\Delta_4)^c\right]\leq 2P\left[\Delta_1^c\right]+P\left[\Delta_3^c\right]+P\left[\Delta_4^c\cap\Delta_1\right]
$$

and Propositions 8, 62 and 63 allow to conclude similarly to the case 1.. $\square$

6.2 Proof of Proposition 61

We only expose the proof for the estimator $\widehat{h}_{\widehat{m}_1}$. The proof of Proposition 61 follows the same line as Proposition 51, let us point out the slight differences. Inequalities (18) and (19), as well as Lemma 51 hold. Hence, for every model m and every $h\in S_m$,

$$
\begin{aligned}
\mathbb{E}\left[\|\widehat{h}_{\widehat{m}_1}-h\|^2_{\overline{F}_Y}1_{\Delta_1\cap\Delta_2\cap\Delta_3}\right] \leq C_1\Bigg\{&\|h-h_m\|^2_{\overline{F}_Y}+\mathbb{E}\left[(pen_1(m)-pen_1(\widehat{m}_1)+2p_1(m,\widehat{m}_1))1_{\Delta_3}\right]+\\
&+\|h\|^2_{\overline{F}_Y}\frac{1}{\overline{F}_0 n}+\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in S_m+S_{m'},\|t\|_{\overline{F}_Y}=1}(\nu_{n,1}(t))^2-\frac{2}{3}p_1(m,m')\right)_+\right]\Bigg\}
\end{aligned}
$$

with

$$
p_1(m,m')=\frac{2B}{5}\frac{K^2}{\overline{F}_0}\frac{D_m+D_{m'}}{n}
$$

The only difference with the proof of Proposition 51 is the upper bound of $\mathbb{E}\left[(pen_1(m)-pen_1(\widehat{m}_1)+2p_1(m,\widehat{m}_1))1_{\Delta_3}\right]$, indeed:

$$
\begin{aligned}
\mathbb{E}\left[(pen_1(m)-pen_1(\widehat{m}_1)+2p_1(m,\widehat{m}_1))1_{\Delta_3}\right] &= \mathbb{E}\left[\left(\frac{B}{\widehat{F}_0}\frac{D_m-D_{\widehat{m}}}{n}+\frac{4B}{5\overline{F}_0}\frac{D_m+D_{\widehat{m}}}{n}\right)1_{\Delta_3}\right]\\
&\leq \mathbb{E}\left[\left(\frac{B}{\widehat{F}_0}\frac{D_m-D_{\widehat{m}}}{n}+\frac{B}{\widehat{F}_0}\frac{D_m+D_{\widehat{m}}}{n}\right)1_{\Delta_3}\right]\\
&= \mathbb{E}\left[\frac{2B}{\widehat{F}_0}\frac{D_m}{n}1_{\Delta_3}\right]\leq\frac{8B}{3\overline{F}_0}\frac{D_m}{n} \quad \square
\end{aligned}
$$

6.3 Proof of Proposition 62

1. If $\alpha_n\leq\overline{F}_0/2$, then

$$
P[\Delta_3^c]=P\left[\left|\frac{1}{n}\sum_{i=1}^n(1_{Y_i\geq 1}-\overline{F}_0)\right|\geq\frac{1}{4}\overline{F}_0\right]=P\left[\left|\frac{1}{n}\sum_{i=1}^n(1_{Y_i\geq 1}-\mathbb{E}[1_{Y_i\geq 1}])\right|\geq\frac{1}{4}\overline{F}_0\right]
$$

   We apply Bernstein Inequality with the parameters $c=1$ and $v=\overline{F}_0$, then $P[\Delta_3^c]\leq 2\exp(-C_1 n\overline{F}_0)$ where C_1 is a numerical constant.
2. We prove that $\Delta_1\cap\Delta_3\subset\Delta_2\cap\Delta_3$. According to (9), on the set $\Delta_1\cap\Delta_3$:

$$
\|t\|_n^2\geq\frac{3}{4}\times\frac{4}{5}\widehat{F}_0\|t\|^2=\frac{3}{5}\widehat{F}_0\|t\|^2
$$

   So $\inf_{t\in S_m,t\neq 0}\frac{\|t\|_n^2}{\|t\|^2}=\min(Sp(\widehat{G}_m))\geq\frac{3\widehat{F}_0}{5}$ thus $\Delta_1\cap\Delta_3\subset\Delta_2\cap\Delta_3$.
   This entails the result 2) from Proposition 62, in fact:

$$
(\Delta_2^c\cap\Delta_3)\cap\Delta_1=\Delta_2^c\cap(\Delta_3\cap\Delta_1)\subset\Delta_2^c\cap(\Delta_3\cap\Delta_2)=\emptyset
$$

   thus $\Delta_2^c\cap\Delta_3\subset\Delta_1^c$. Besides we have immediatly $\Delta_2^c\cap\Delta_3\subset\Delta_3$, so $\Delta_2^c\cap\Delta_3\subset\Delta_1^c\cap\Delta_3$. $\square$

6.4 Proof of Proposition 63

Let x_0 and $\widehat{x}_0$ be in A such that:

$$\|h\|_\infty = h(x_0), \quad \|\widehat{h}_D\|_\infty = \widehat{h}_D(\widehat{x}_0)$$

Then,

$$\widehat{\nu}_n - \nu \le (\widehat{h}_D - h)(\widehat{x}_0) = (\widehat{h}_D - h_D)(\widehat{x}_0) + (h_D - h)(\widehat{x}_0) \le \sqrt{D} \sup_{j=1,\dots,D} |\widehat{a}_j - a_j| + \|h - h_D\|_\infty$$

Similarly,

$$\nu - \widehat{\nu}_n \le (h - \widehat{h}_D)(x_0) \le (h - h_D)(x_0) + (h_D - \widehat{h}_D)(x_0) \le \|h - h_D\|_\infty + \sqrt{D} \sup_{j=1,\dots,D} |\widehat{a}_j - a_j|$$

Hence $|\nu - \widehat{\nu}_n| \le \|h - h_D\|_\infty + \sqrt{D} \sup_{j=1,\dots,D} |\widehat{a}_j - a_j|$. According to Assumption (15),

$$\begin{aligned} P[\Delta_4^c] &\le P\left[\|h - h_D\|_\infty + \sqrt{D} \sup_{j=1,\dots,D} |\widehat{a}_j - a_j| \ge \frac{\nu}{4}\right] \\ &\le P\left[\sqrt{D} \sup_{j=1,\dots,D} |\widehat{a}_j - a_j| \ge \frac{\nu}{8}] \le \sum_{j=1}^{D} P[\sqrt{D}|\widehat{a}_j - a_j| \ge \frac{\nu}{8}\right] \end{aligned}$$

Moreover, for every $j = 1, \dots, D$,

$$\begin{aligned} \sqrt{D}(\widehat{a}_j - a_j) &= \sqrt{D}\left[\frac{(1/n)\sum_{i=1}^n \delta_i \varphi_j^D(Y_i)}{\|\varphi_j^D\|_n^2} - \frac{\int_A \varphi_j^D(x)h(x)\overline{F}_Y(x)dx}{\|\varphi_j^D\|_{\overline{F}_Y(x)}^2}\right] \\ &= \frac{\sqrt{D}}{\|\varphi_j^D\|_n^2} \frac{1}{n} \sum_{i=1}^n \left[\delta_i \varphi_j^D(Y_i) - \int_A \varphi_j^D(x)h(x)\overline{F}_Y(x)dx\right] + \\ &\quad \sqrt{D} \int_A \varphi_j^D(x)h(x)\overline{F}_Y(x)dx \left[\frac{1}{\|\varphi_j^D\|_n^2} - \frac{1}{\|\varphi_j^D\|_{\overline{F}_Y}^2}\right] \end{aligned}$$

Besides, on the set Δ_1,

$$\|\varphi_j^D\|_n^2 \ge \frac{3}{4}\|\varphi_j^D\|_{\overline{F}_Y}^2 = \frac{3}{4}\int_{(j-1)/D}^{j/D} D\overline{F}_Y(x)dx \ge \frac{3\overline{F}_0}{4}$$

and

$$\left|\int_A \varphi_j^D(x)h(x)\overline{F}_Y(x)dx\right| \le \|h\|_{\overline{F}_Y}\|\varphi_j^D\|_{\overline{F}_Y} \le \|h\|_{\overline{F}_Y}$$

Hence:

$$\begin{aligned} \sqrt{D}|\widehat{a}_j - a_j|1_{\Delta_1} &\le \frac{4\sqrt{D}}{3\overline{F}_0}\left|\frac{1}{n}\sum_{i=1}^n [\delta_i \varphi_j^D(Y_i) - \int_A \varphi_j^D(x)(h)\overline{F}_Y(x)dx\right| + \sqrt{D}\|h\|_{\overline{F}_Y}\left|\frac{\|\varphi_j^D\|_{\overline{F}_Y}^2 - \|\varphi_j^D\|_n^2}{\|\varphi_j^D\|_{\overline{F}_Y}^2\|\varphi_j^D\|_n^2}\right| \\ &\le \frac{4\sqrt{D}}{3\overline{F}_0}\left|\frac{1}{n}\sum_{i=1}^n (\delta_i \varphi_j^D(Y_i) - \mathbb{E}[\delta_i \varphi_j^D(Y_i)])\right| + \sqrt{D}\|h\|_{\overline{F}_Y}\frac{4^2}{3^2\overline{F}_0^2}\left|\|\varphi_j^D\|_{\overline{F}_Y}^2 - \|\varphi_j^D\|_n^2\right| \end{aligned}$$

Thus:

$$\begin{aligned} P[\Delta_4^c \cap \Delta_1] &\le \sum_{j=1}^D P\left[\frac{4\sqrt{D}}{3\overline{F}_0}\left|\frac{1}{n}\sum_{i=1}^n (\delta_i \varphi_j^D(Y_i) - \mathbb{E}[\delta_i \varphi_j^D(Y_i)])\right| \ge \frac{\nu}{16}\right] + \\ &\quad \sum_{j=1}^D P\left[\sqrt{D}\|h\|_{\overline{F}_Y}\frac{4^2}{3^2\overline{F}_0^2}\left|\|\varphi_j^D\|_{\overline{F}_Y}^2 - \|\varphi_j^D\|_n^2\right| \ge \frac{\nu}{16}\right] \\ &:= \sum_{j=1}^D (P_{1,j} + P_{2,j}) \end{aligned} \tag{27}$$

$P_{1,j}$ and $P_{2,j}$ are upper bounded with Bernstein Inequality. For $P_{1,j}$, the parameters b and v are the following:

$$\mathbb{E}[\delta_i^2(\varphi_j^D)^2(Y_i)] = \int_A (\varphi_j^D)^2(x)h(x)\overline{F}_Y(x)dx \leq \|h\|_{\infty,A} := v, \quad \|\delta_i\varphi_j^D(Y_i)\|_\infty \leq \sqrt{D} := c$$

Hence, for every $j \in \{1,\ldots,D\}$,

$$P_{1,j} \leq 2\exp(-C\frac{n}{D}) \tag{28}$$

where C depends on $(\nu, \|h\|_{\infty,A}, \overline{F}_0)$. Let us upper bound $P_{2,j}$. For every $j \in \{1,\ldots,D\}$,

$$P_{2,j} = P\left[\frac{4^2\sqrt{D}\|h\|_{\overline{F}_Y}}{3^2\overline{F}_0^2}\left|\frac{1}{n}\sum_{i=1}^n(\int_A(\varphi_j^D)^2(x)1_{Y_i\geq x}dx - \mathbb{E}[\int_A(\varphi_j^D)^2(x)1_{Y_i\geq x}dx])\right| \geq \frac{\nu}{16}\right]$$

and

$$\mathbb{E}\left[\left(\int_A(\varphi_j^D)^2(x)1_{Y_i\geq x}dx\right)^2\right] \leq 1 := v, \quad \left\|\int_A(\varphi_j^D)^2(x)1_{Y_i\geq x}dx\right\|_\infty \leq 1 := c$$

Thus, with Bernstein Inequality we obtain

$$P_2 \leq 2\exp\left(-C\frac{n}{D}\right) \tag{29}$$

where C depends on $(\nu, \|h\|_{\overline{F}_Y}, \overline{F}_0)$. Now inequalities (27), (28) and (29) entails the result of Proposition 63. □

7 Appendix

The following Inequality, called Bernstein Inequality, is proved in this form in Birgé and Massart (1998), p366, Lemma 8.

Theorem 3 *Let $(X_1,\ldots,X_n)$ be independent random variables. Let us suppose that:*

$$\frac{1}{n}\sum_{i=1}^n\mathbb{E}[X_i^2] \leq v, \quad \frac{1}{n}\sum_{i=1}^n\mathbb{E}[(X_i)_+^l] \leq \frac{l!}{2}\times v\times c^{l-2}$$

for every $l \geq 2$. Let $S = \frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_i]$.

1) For every $\epsilon > 0$:

$$P[S \geq \sqrt{2vx} + cx] \leq \exp(-nx), \quad P[|S| \geq \sqrt{2vx} + cx] \leq 2\exp(-nx).$$

2) Similarly, for every $\epsilon > 0$:

$$P[S \geq \epsilon] \leq \exp\left(-\frac{n\epsilon^2}{2(v+c\epsilon)}\right), \quad P[|S| \geq \epsilon] \leq 2\exp\left(-\frac{n\epsilon^2}{2(v+c\epsilon)}\right).$$

Theorem 4 *Let $\mathcal{F}$ a set of uniformly bounded functions, which have a countable dense for the infinite norm subspace. Let $(X_1,\ldots,X_n)$ be independant random variables and:*

$$Z = \sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n[f(X_i) - \mathbb{E}(f(X_i))]\right|$$

Let consider b, v and $\mathbb{H}$ such that:

$$b \geq \sup_{f\in\mathcal{F}}\|f\|_\infty, \quad v \geq \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n Var(f(X_i))$$

$$\mathbb{H} \geq \mathbb{E}(Z)$$

Then for every $\theta > 1$, there exist numerical constants $\overline{C}$, $\overline{C}'$, $\overline{\kappa}$, $\overline{\kappa}'$ such that:

$$\mathbb{E}(Z^2 - \theta\mathbb{H}^2)_+ \leq \overline{C}\frac{v}{n}\exp\left(-\overline{\kappa}\frac{n\mathbb{H}^2}{v}\right) + \overline{C}'\frac{b^2}{n^2}\exp\left(-\overline{\kappa}'\frac{n\mathbb{H}}{b}\right)$$

The above version of Talagrand Inequality is enunciated for example in Lacour (2008) (Section 6, Lemma 5).

References

Antoniadis A, Grégoire G, Nason G (1999) Density hazard rate estimation for right-censored data by using wavelet methods. J R Statist Soc B 61(1):63–84

Baraud Y (2002) Model selection for regression on a random design. ESAIM Probab Statist 6:127–146

Birgé L, Massart P (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli 4(3):329–375

Comte F, Brunel E (2005) Penalised contrast estimation of density and hazard rate with censored data. Sankhya 67(3):441–475

Comte F, Brunel E (2008) Adaptive estimation of hazard rate with censored data. Comm Statist Theory Methods 37(8-10):1284–1305

Huber C, MacGibbon B (2004) Lower bounds for estimating a hazard. Handbook of Statistics, Elsevier, Amsterdam 23:209–226

Kaplan E, Meier P (1958) Non parametric estimation from incomplete observations. J Amer Statist Assoc 53:457–481

Lacour C (2008) Adaptive estimation of the transition density of a particular hidden markov chain. J Multivariate Anal 99(5):787–814

Muller H, Wang J (1994) Hazard rate estimation under random censoring with varying kernels and bandwidth. Biometrics 50(1):61–76

Nelson W (1972) Theory and applications of hazard plotting for censored failure data. Technometrics 14(4):945–966

Patil P (1993) On the least squares cross validation bandwidth in hazard rate estimation. Ann Statist 21(4):1792–1810

Reynaud-Bouret P (2006) Penalized projection estimators of the Aalen multiple intensity. Bernoulli 12(4):633–661

Tanner M, Wong W (1983) The estimation of the hazard function from randomly censored data by the kernel method. Ann Statist 11(3):989–993

Yandell B (1983) Nonparametric inference for rates with censored survival data. Ann Satist 11(4):1119–1135